

ITBW41 Predictive Analytics Project

Project Proposal

Healthcare Predictions

| Submitted by: | Shermaine Peh (213276Q)\*  Koh Shi Min (201357X)  Zhang Xiang (210896X)  Rawtbhik (210902G) |
| --- | --- |
| Module group: | ITBW41 - 01 |
| Tutor: | Mr Law Chee Yong |
| Report Last Amended: | 15th July 2022 |

**Table of Contents**

[**Introduction & Background**](#_yqm85j5p5k7i) **3**

[Introduction](#_txmat5o5ph5n) 3

[Background](#_77e726vpj7m0) 3

[**Business Scenario & Objective**](#_o5w8l6ipl6hm) **4**

[Scenario & Objective](#_61otfxra3xoi) 4

[Empathy](#_aeyvvzjx3it5) 4

[Transdisciplinary Thinking & Pitching](#_hiltqrkyg2bf) 5

[**Main & Sub-tasks**](#_9zdp6kdzi6b6) **6**

[**Task Allocation & Project Schedule**](#_uwb3yqwysmzy) **7**

# **Introduction & Background**

## ***Introduction***

Our company, SSRZ, focuses on a mission to provide clients with predictive modelling, using Artificial Intelligence and Machine Learning. Having worked with many clients over the years, our company is confident that with the help of **professional input**, we will offer our finest model and recommendations to address any problems.

## ***Background***

Using our knowledge of Supervised Learning techniques, we aim to provide categorical data-driven predictions and accuracy in the healthcare industry with labelled data, helping to catch up with the growing number of patients and lack of manpower, while also allowing for early detection.

We will use the Supervised Machine Learning Techniques as follows:

1. Correlation Matrix
2. Logistic Regression
3. Decision Tree & Random Forest
4. Support Vector Machine
5. Neural Network

By utilising software programs like **PowerBI, Tableau, KNIME, SAS Viya, and Figma** to help meet our business goals, our company will provide diverse ideas, knowledge, and specific **industry experience**. Additionally, our proposed timeline can be found in our [schedule and task allocation](#_uwb3yqwysmzy).

# **Business Scenario & Objective**

## ***Scenario & Objective***

In the present situation, our main objective is to propose a project to **develop categorical predictive models of 85% accuracy** to detect **common health problems in our global population and the factors** contributing to those health issues. As health problems are on the **rise**, our team believes that **early detection and response** are essential and a **vital** part of providing healthcare services with **accuracy and effectiveness**. Additionally, these predictions will help **pinpoint factors associated with major health problems**, so that potential patients will be **aware** of them and take **preventive** measures to **avoid** becoming victims of such problems.

Due to a **shortage** of healthcare professionals, it will also be difficult to provide **high-quality and accurate** care to every patient. Hence, we are confident that our machine learning techniques will be able to detect health problems more **accurately** through algorithms and computer analysis, thereby **reducing healthcare stress, delayed detection and human error.**

## ***Empathy***

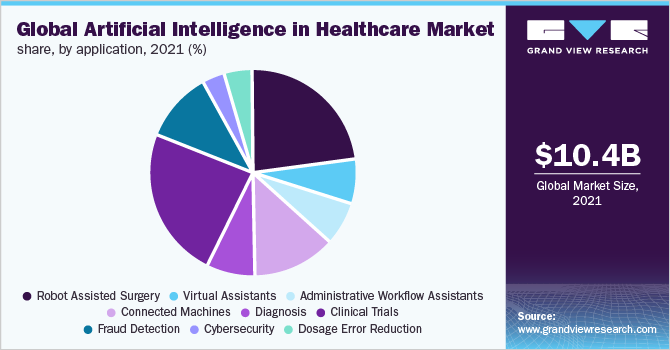
**Engage** - During our interviews with the two doctors mentioned above, we found that **strokes** and **diabetes** are on the **rise**, requiring **prompt** and **accurate** detection.

**Observe** - Looking at our current healthcare conditions, especially with infectious diseases on the rise, we see that the healthcare industry is **severely understaffed**, so they will need help detecting common health problems to **reduce** their workload, checking and recommending each patient.

## 

## ***Transdisciplinary Thinking & Pitching***

In order to gain a better understanding of the problem, our team took the early initiative to consult two **professionals**, Dr Wee Chee Keong and Dr Alvin Ng. Through their input, we learned that the two main and increasing diseases faced by the current population are stroke and diabetes, which we will help **detect** and gain better **insights** into throughout this project. In addition, we learned that many of the health problems faced by our current population are related to **age, hypertension, diet, and whether or not people smoke**. Using this information, our team will then gather datasets on the following problems and factors, before providing a complete **prediction and recommendation.**



With the graph, AI has proven **invaluable** in many different areas of healthcare. However, **error reduction and early detection** have not yet had a significant impact. Hence, our team aims to use our **predictive** models to make a bigger **impact** by contributing to **early detection and accuracy** in the healthcare industry.

## 

# **Main & Sub-tasks**

***Main Task***

* Determine the **important** factors that contribute to an individual getting **common health problems**
  + Data Collection
  + Data Understanding
  + Data Cleaning / Pre-processing
  + Data Modelling
  + Report Writing

***Sub-Task***

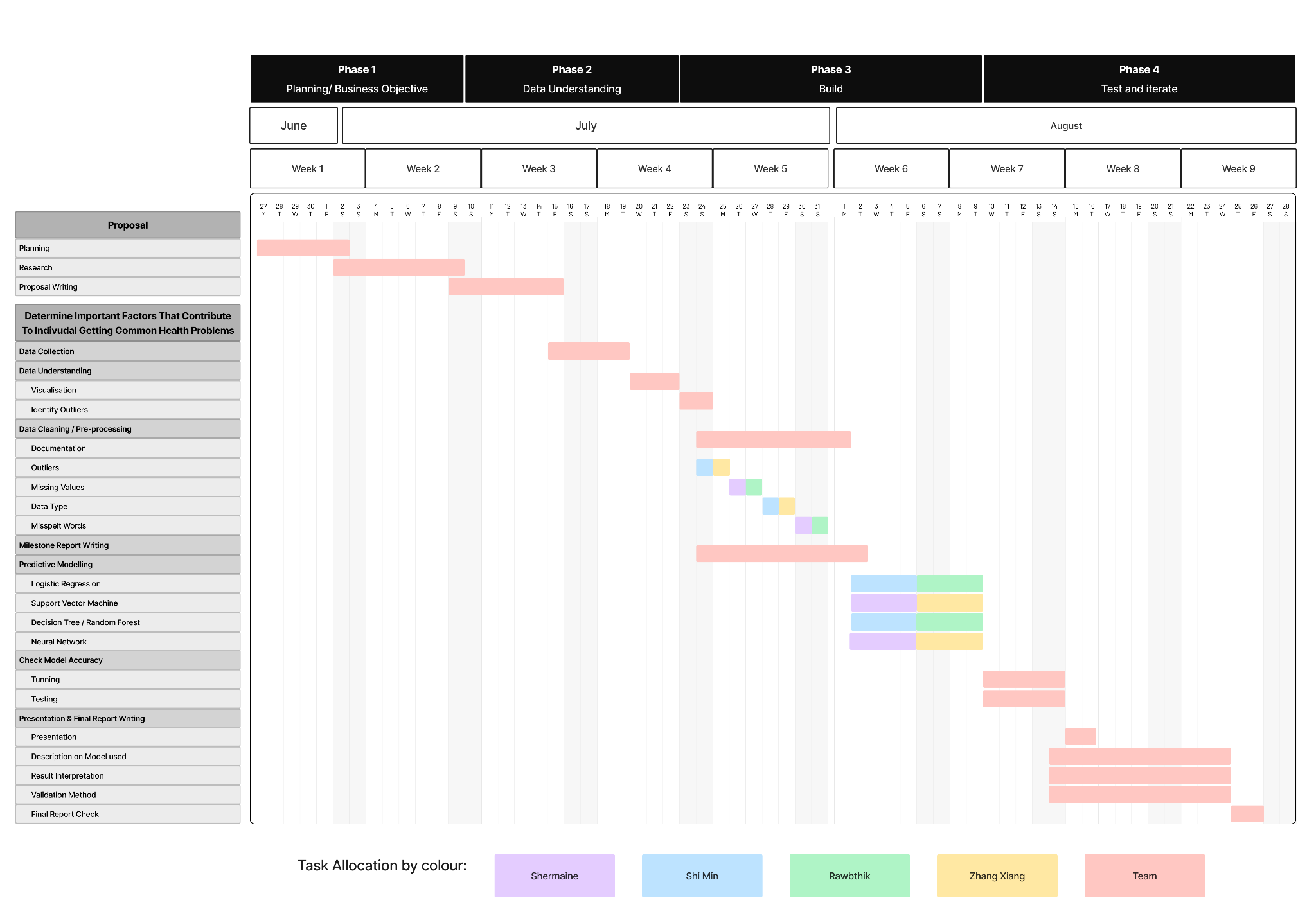
* Provide final predictive of at least **85% accuracy**
* Identify business objectives
  + Select our **datasets** to use
  + Determine which **industry** we want to work with
  + Consult industry **professionals**
* Prepare data for modelling
  + Data **visualisation** and **understanding** using software tools mentioned [here](#derlme3vkcid)
  + Select which variables to use
* Data modelling
  + Focusing on logistic regression, support vector machine, decision tree/ random forest, neural network and with the help of **SAS Viya**, we will train, test and tune our models to provide our **most accurate and effective** model for **recommendations**
* Report writing
  + **Document** down our process throughout this project
  + **Combine** our different models and documentations
  + Set and meet **timelines** (Gantt Chart)
* Final presentation
  + **Finalising** and selecting **best** model
  + Prepare slides and rehearsal

# **Task Allocation & Project Schedule**

***Task Allocations***

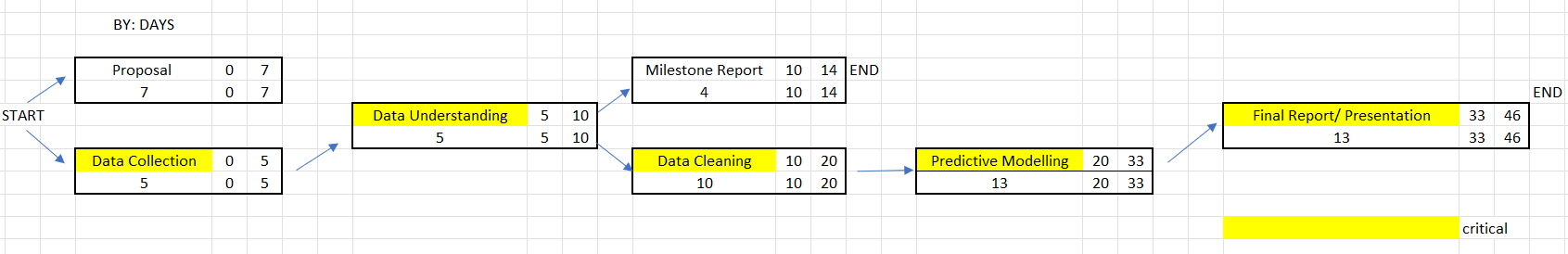
| **Tasks** | **Shermaine** | **Shi Min** | **ZhangXiang** | **Rawtbhik** |
| --- | --- | --- | --- | --- |
| Data Collection | ✔️ | ✔️ | ✔️ | ✔️ |
| Data Understanding |  |  |  |  |
| Visualisation | ✔️ | ✔️ | ✔️ | ✔️ |
| Outliers Detection | ✔️ | ✔️ | ✔️ | ✔️ |
| Data Cleaning / Pre-processing |  |  |  |  |
| Outliers |  | ✔️ | ✔️ |  |
| Missing Values | ✔️ |  |  | ✔️ |
| Data Type |  | ✔️ | ✔️ |  |
| Misspelt words | ✔️ |  |  | ✔️ |
| Predictive Modelling |  |  |  |  |
| Logistic Regression |  | ✔️ |  | ✔️ |
| Support Vector Machine | ✔️ |  | ✔️ |  |
| Decision Tree/Random Forest |  | ✔️ |  | ✔️ |
| Neural Network | ✔️ |  | ✔️ |  |

***Project Schedule (Gantt Chart)***

******

***CRITICAL TASK***

The below critical task aims to serve as a baseline to show a quick overview of how **important** and long each main task will take. With the **critical path**, we can also better understand that every task needs to be **considered** along the way, so as to **not delay or hold up** future tasks.

******